Unbiased AI Analysis of Full-Scale Mass Spectrometry Data Identifies Disease-Relevant Treatment-Response Signals

Key Point

Magellan's platform enables analysis of the full output of mass spectrometry—nearly one million datapoints per sample in this study—without pre-filtering or bias. From this comprehensive dataset, the platform identified approximately 180 highly relevant features that cleanly distinguish pre-treatment patients from post-response patients.

Among the 180, the features identified as peptides were further analyzed via MS/MS. Their identities were found to be associated with disease-relevant biology, despite Magellan being blinded to study indication or therapeutic mechanism of action beforehand.

These results demonstrate that Magellan's software can extract biologically meaningful, non-random signals from complete and large mass spectrometry datasets. This capability enables a fundamentally new and unbiased approach to biomarker discovery, patient stratification, and understanding the molecular basis of disease and its treatment.

Background

Blood is a dynamic and information-rich substrate that reflects biological activity across tissues. Mass spectrometry can capture this complexity at scale, detecting hundreds of thousands to millions of molecular features per sample. However, in conventional workflows, most of this data is discarded because of software processing limitations. Analyses are typically restricted to predefined targets or heavily filtered datasets, introducing bias at the outset and limiting discovery.

Magellan's platform addresses this limitation by retaining and analyzing the full mass spectrometry output. Rather than pre-selecting features, all detectable molecular signals are considered equally. Magellan's software then identifies which signals are relevant to the biological or clinical question. This unbiased and comprehensive analytical framework enables extraction of meaningful biological insight from complex and large datasets.

Methods and Results

This study analyzed paired plasma samples collected before treatment and after successful clinical response from 10 patients (20 total samples). Magellan was blinded to trial indication, therapeutic mechanism, and response status beyond sample pairing.

Samples were processed to release small peptides and other biomarkers from the protein-bound fraction while removing high-abundance, low-information proteins. The resulting sample was enriched for low-abundance species that reflect ongoing biological processes.

LC-MS was used to detect and measure molecular species found in each sample. Each run produced approximately 945,000 digitized datapoints after excluding those containing no signal. Each datapoint represents a measurable molecular feature in plasma. Unlike conventional approaches, all datapoints were retained and analyzed.

From this full dataset, Magellan's software identified ~180 datapoints with a Decision Value greater than 1, indicating perfect, non-overlapping separation of pre-treatment and post-response expression. Magellan's machine learning analysis confirmed that this level of separation is strongly connected to underlying biology of treatment response and cannot be explained by random noise.

Nearly all identified features showed increased expression following successful treatment, consistent with a coordinated biological response.

The 180 features included both peptide and non-peptide signals. The sponsor prioritized peptide identification for follow-up. Targeted LC-MS/MS analysis with low-energy fragmentation was applied to the regions corresponding to these features. Ten peptides were successfully identified and mapped to their parent proteins using PEAKS software.

These proteins are associated with biological processes directly relevant to the disease and treatment response, including immune activity, fibrosis, and epithelial-mesenchymal transition. These relationships became clear only after unblinding. Their molecular functions included regulation of protein degradation and chromatin, implicating pathways not

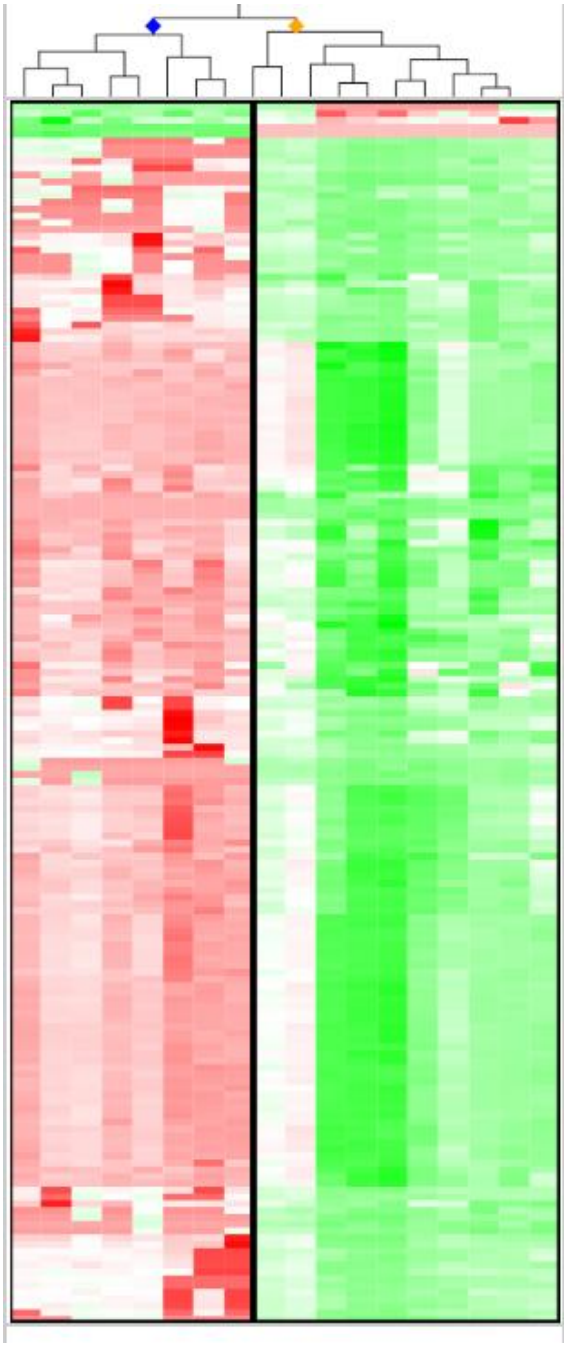

Figure 1. Novel biomarker expression patterns.

Relative expression of the 180 biomarkers associated with treatment response. Columns under the blue and yellow diamonds indicate the expression of datapoints for samples taken pre- treatment or post-response, respectively. Note that the analytics indicates altered treatment response for specific patients (rightmost 2 post-treatment samples).

previously associated with the drug's mechanism of action.

Conclusion

This study demonstrates that Magellan's platform can capture and analyze entire mass spectrometry datasets without bias to identify signals connected to a clinical parameter.

Starting from nearly one million datapoints per sample, the platform detected ~180 features connected to successful treatment. These features were not only statistically robust but also biologically meaningful, mapping to disease-relevant processes despite the analysis being blinded to indication and drug mechanism.

By combining a data-rich biological source (plasma), the full analytical depth of mass spectrometry, and software capable of processing all resulting data, Magellan extracts biological information that has historically been inaccessible.

The results show that Magellan can identify markers that both stratify patients and reflect underlying disease biology. This establishes a new, unbiased path to biomarker discovery and mechanistic insight in human studies.

info@mag-bio.com